Chapter 14

SOME APPROACHES TO COMPUTERIZED MEDICAL DIAGNOSIS

T. ALLAN PRYOR
HOMER R. WARNER

THE CLASSIFICATION OF A patient into a disease category having therapeutic and prognostic usefulness may be viewed as a stepwise procedure. One step represents the reduction of raw data into a set of parameters. A second step might deduce from these parameters a classification of these test values (i.e. ECG waveform), and a final step arrive at a diagnosis from the set of test classifications. As an example of steps 1 and 2, we will present a discussion of some approaches to ECG classification. Step 3 will be illustrated by examples of patient diagnosis using both binary and distributed information.

During the past six years, four methods of classifying electrocardiograms using the computer have been tried at the Latter-day Saints Hospital. In each case, a modified Frank lead system was used to generate an X, Y, Z lead. The first method used Markov sequences. The present state of the sequence was defined as the last N sampled points of a waveform. From this state the next state or M points on the waveform was predicted. The sequences were constructed by sampling waveforms from known classifications and generating frequency distributions for the changes from one state to the next. With this information the final predictive matrices for each category could be generated. The number of possible states is l^{dn}, where l is the number of values each point may assume and d is the dimension of the waveform and n is the number of points back in time (Fig. 14–1). Two waveforms were used. The first was the spacial velocity generated by the X, Y, and Z values, i.e. $S.V. = \sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}$. In this case d is equal to 1. Hence, the

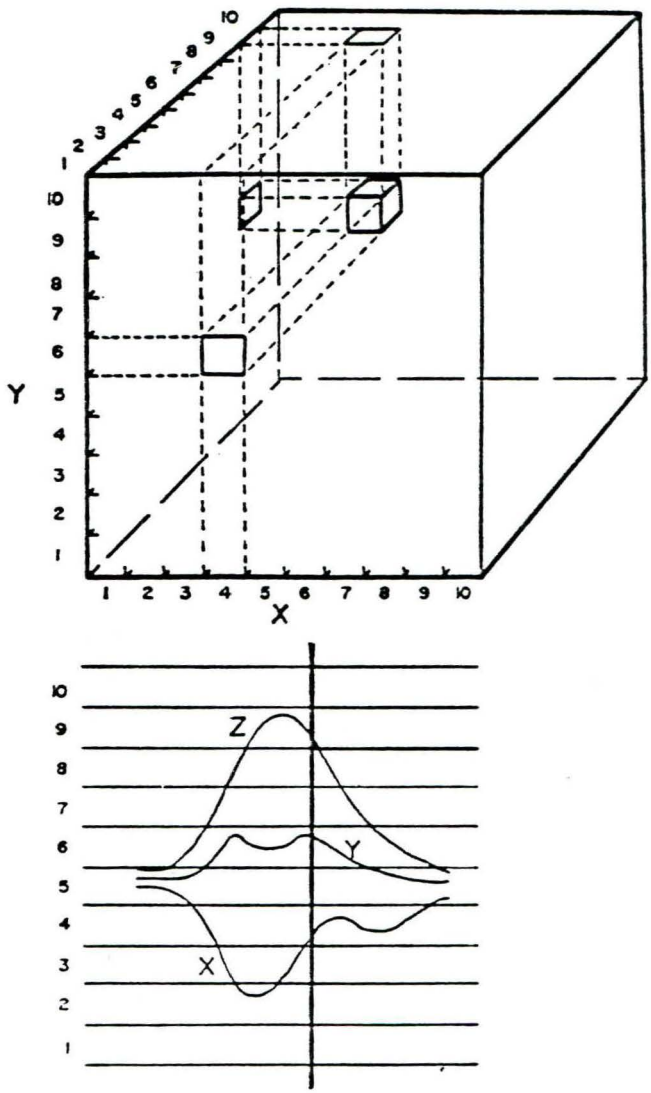

Figure 11–1. Three dimensional histogram for ECG interpretation.

state matrix could contain several points back in time and fit within available computer memory. In the second application of the predictive histogram the actual sampled X, Y and Z values were used. Here, d is equal to 3 and the values of X, Y and Z were digitized into only 10 possible values. With just one point back-

ward in time, the matrix size is 1000 cells and for two points backwards in time, the size of the matrix increases to one million cells. Reductions, however, could be made in the size of the matrix since the values at any one point cannot vary over the complete range of values due to the continuous nature of the waveform. None the less, the problem remains the same, that is, the size of the matrices. To classify a waveform, the points of the waveform were successively used to predict a point on a theoretical curve. This was done using the state information from each of the matrices of known classification. The theoretical curves could then be correlated against the actual to determine the class in which the unknown belonged. The major problem with this technique results from the size of the matrices necessary to store the state information. That is, if the number of points backwards in time was small, say one point, then the changes predicted were only local changes with time. This caused the generation of similar waveforms from each state matrix giving no discrimination between the classes. If more points back in time were used the size of the matrices and time for computation became excessive.

The second method continued the use of histograms. However, in this case, rather than being predictive histograms, they were in fact three-dimensional amplitude histograms. Again the known waveforms were sampled and each sample in time was used to develop the histogram for that class. To diagnose a waveform it was sampled, a histogram generated from this waveform and this histogram was correlated against the various class histograms. A probability was assigned to each class by the following equation:

$$P_i = \frac{C_i}{\Sigma C_j} \qquad (1)$$

C_j is the correlation of the unknown histogram to that of the j in class. A wide variety of wave shapes in one class resulted in a smoothing of the histogram. Hence, if all possible shapes from one classification were included in the generation of a single histogram the resulting histogram would correlate equally well against any unknown waveform. As a result, a method of adaptive generation of the histogram was tried. Here, as new cases were con-

sidered, a correlation against its histogram and the known histograms were determined. If the correlation was greater than .80 the new waveform was added into the histogram of that particular class. Unfortunately, this criteria resulted in the generation of an arbitrarily large number of different classifications and histograms.

At this point the histogram technique was dropped and linear discriminant analysis was tried. The program used was the standard Bi-Med discriminant analysis program contained in the Bi-Med program series. This program performed the ECG classification from a set of parameters measured from specific components of the sampled ECG waveforms. Thus, for the first time in the project it became necessary to derive parameters from the original data before proceeding to step 2 as defined in the introduction.

With this program the question, of course, arises of which parameters are most significant in discriminating among the various classifications. Again, variations within the classes presented a problem. Since many of the original parameters did not discriminate between certain classes, they tended to lessen the discrimination among the classes. A technique of weighting various parameters under certain classifications was considered. This, in fact, results precisely in the tree logic which is presently being used. Statistics were generated on the parameters from the discriminant function program to determine individually their discriminatory value among the classes. With this information available it was then possible to develop the present program which follows a tree structured diagnostic logic where the waveform is checked against certain parameters, and if the test is positive the waveform is classified into that category, and if not, a new category is checked by looking only at those parameters important to that particular classification. This program is being used routinely on approximately 50 patients per day at two hospitals in Salt Lake City, and represents one source of input data for a more general diagnostic program (step 3).[1] Table 14–I gives the results of a study using the program.

The original work in medical diagnosis done here was in 1961 and applied Bayes' Theorem to the diagnosis of congenital heart

TABLE 14–I
RESULTS OF COMPUTER ANALYSIS OF ELECTROCARDIOGRAM

	N	*AI*	*II*	*LI*	*LVH*	*RVH*	*LBBB*	*RBBB*
N	58	1						2
AI		20						
II	1		7					
LI				2				
LVH	1	1			6			
RVH						1		
LBBB							5	
RBBB								4

disease.[2,3] Thirty-three diseases were considered and 50 symptoms. Equation 2 gives the form of Bayes' Theorem used.

$$P_{D_i|S_1S_2\cdots S_n} = \frac{P_{D_i} \cdot \prod_{j=1}^{n} P_{S_j|D_i}}{\sum_{k=1}^{e} P_{D_k} \cdot \prod_{j=1}^{n} P_{S_j|D_k}} \qquad (2)$$

This form of the equation assumes that the symptoms are independent within a given patient; that is, if a patient has symptom 1 he is no more likely to have symptom 2 than if he did not have symptom 1. The equation also assumes that the patient will have one and only one of the diseases in the set. This means that any possible combinations of the diseases that can occur in a single patient must be included as a separate disease. In spite of the fact

that many of the symptoms indeed are independent, the method still proves useful since one of two courses can be taken. First, if the two symptoms prove to be highly dependent upon another they may be included as a single symptom. For instance, cyanosis and clubbing of the fingers occur together almost invariably if the cyanosis is severe. Thus, these two symptoms can be made one under the term "severe cyanosis." On the other hand, if the dependence is not a strong one, it may be better to include two not quite independent symptoms as though they were independent, since the amount of information contributed by each is significantly greater than the loss of information incurred if one of the symptoms is ignored. This can be tested empirically by examining the diagnostic performance of the system with and without the symptom in question.

A program was developed which allowed this system to be run from a time-shared computer station at which the symptoms were entered in response to a series of lists of symptoms on the oscilloscope. The physician simply presses the appropriate keys at the console to indicate which of the symptoms presented were actually present in the patient under examination. Those attributes were then used in making the diagnosis. The ability of this computer program to complete the logical process and arrive at the correct diagnosis was tested against the physician who collected the initital data; thus both physician and computer used exactly the same data upon which to make their deductions. The physician was asked to enter his differential diagnosis on the back of the entry data sheet and list by each entry a probability. A system for scoring the physician's performance against the computer was derived using the product of two terms. The first term represents the fraction of times that the computer or physician gave the patient the correct diagnosis with a probability of at least 1 percent. The second term reflects his confidence in the correct diagnosis when he made it and consists of the probability that he assigned to the correct diagnosis. In over 200 cases diagnosed by both physician and computer, the computer program clearly performed better than the physician in all cases except one doctor who was an experienced pediatric cardiologist. He and the computer performed equally well from the same information.

In 1966 another approach to the diagnosis of congenital heart disease was tried in fulfillment of a master of science degree in the Computer Science Department of the University of Utah.[4] This approach used discriminant functions derived from the assumption of multivariant normal distribution of the symptoms. The discriminant functions are given by the following equation:

$$g_i(x) = C_i - \tfrac{1}{2}\ (X\text{-}M_i)^t\, \Sigma^{-1}\, (X\text{-}M_i)\ \ i=1, \ldots R \qquad (3)$$

where

$$C_i = \log p(i) - \tfrac{1}{2} \log|\Sigma_i| \qquad (4)$$

This function defines the disease which maximizes $g_i(x)$ as the most probable disease that these symptoms represent. The use of this equation requires the generation of the covariant matrix (Σ_i) and the mean matrix (M_i) for each disease. These matrices are generated using selected cases of known diseases. In the test which was run, 508 cases were available. Three hundred of these were used as learning observations to generate the covariant and mean matrices for each disease. In those diseases where there were less than 20 cases, the same observations were used over again until at least eight observations were available for each disease considered. Of the 33 diseases which were to be considered, only 28 had any data available. Table 14–II shows a comparison on selected diseases between the multivariate program and the Bayesian diagnostic program. As can be seen from the table there were in some instances improvements using the multivariate analysis, and in some instances the results were not as good. Ex-

TABLE 14–II
COMPARISON OF BAYES' PROGRAM VS MULTIVARIATE PROGRAM
(BINARY SYMPTOMS)

Disease	*Number of Cases*	*Number Correct (MULTI)*	*Number Correct (BAYES)*	*Improvement in % of Correct Cases*
Normal	93	68	70	1
Atrial Septal Defect	85	57	65	−1
Atrio-ventricular Communis	18	15	14	5
Pulmonary Stenosis Gradiant < 40 mm Hg	34	5	11	−12

amining the overall performance of the program, the multivariate program gave the correct diagnosis 80 percent of the time and Bayesian program 60 percent. Whether this indicates a significant improvement is questionable. It is true, however, that the multivariate program as it presently stands is certainly no worse than the Bayesian approach, the major disadvantage being in the time required to perform the analysis. Since no significant improvement in the majority of diseases were shown using this method, the techniques here have not been continued on this application. This decision is based primarily on the simplicity of the Bayesian approach, the amount of data to be stored and updated in the case of the multivariate analysis and the time required to perform the analysis. The time necessary for one diagnosis of the multivariate is two minutes, whereas for the Bayesian equation the time was in the order of milliseconds. Thus, the Bayesian approach has the advantage of easy implementation as an on-line program than can be run from any terminal on the system.

Since the majority of information available to the physiologist or physician comes not in terms of discrete data as used in the congenital heart diagnosis program, but in terms of continuous data, it was necessary to develop a technique for including within any diagnostic program data of a continuous nature. Examples of this might be pressures within a heart chamber in determining acquired heart disease, volume of packed red cells in determining polycythemic states, etc. A technique has been developed for combining both discrete and continuous data in a Bayesian approach to diagnosis. In order to do this, data from each symptom in a given disease was fit to probability density function. In particular, a lag normal distribution was used as the most simple distribution which fit the continuous data. This distribution was originally studied on an analog computer to describe the distribution of transit time of dye particles in the circulation. The function was generated by taking a normal density curve and running the output through a first order lag circuit. The equation of the lag normal curve is as follows:

$$g(x)=f(x)-\tau g'(x) \quad \text{where } f(x)=\frac{1}{\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2} \qquad (5)$$

This equation can be shown to be a density function under the condition that tau be positive. With this restriction, only curves which are skewed to the left are generated. However, in practice a transformation may be applied to the original data before attempting to fit the data to this curve if the data is skewed to the right. This is done by inverting the curve about its mean value, i.e. subtracting each value from twice the mean value. This equation also has the property that the mean is equal to $\mu+\tau$ where μ is the mean of the normal density function in the equation. The variance of the curve is equal to $\sigma^2+\tau^2$ where again σ^2 is a variance of the normal density function. The data which has been collected for a particular symptom in a given disease is fit to a particular lag normal distribution with μ, σ, τ, being determined for that fit. These values are then stored as part of the a priori probability matrix to be used subsequently in the conditional probability calculations. This is done for each symptom in each disease under consideration. Fig. 14–2 shows a fit of a symptom.

For a given diagnosis then, an actual probability for that symp-

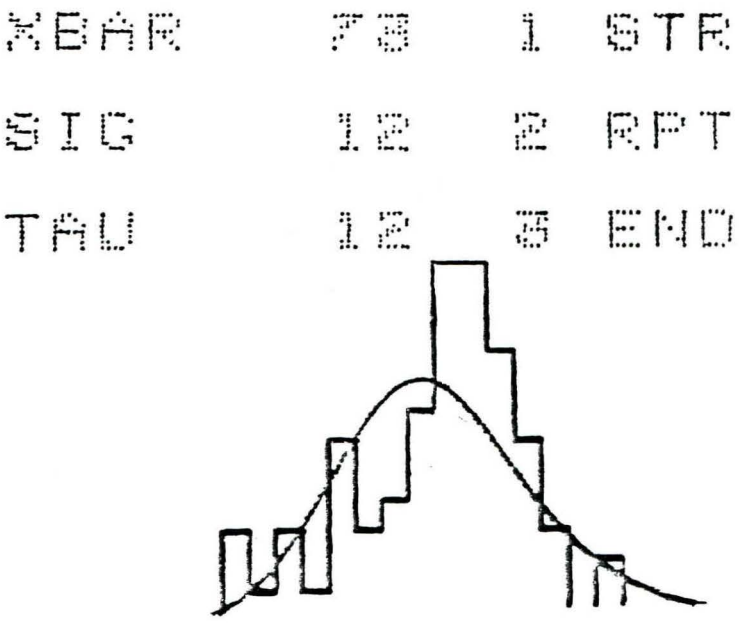

Figure 14–2. Fit of a lag normal curve to a histogram.

tom (test value) given a particular disease is calculated from the stored parameters for the symptoms and diseases. This is done by integrating under the theoretical distribution curve about the measured value of the symptom. Since for each symptom and for each disease there is a probability curve, the probability of finding any particular symptom (test) value in any disease can be determined. This term is then the $P_{s}/_{(o)}$ for that symptom used in the Bayesian formula.

The first application of this program was in 1967 for diagnosis of rheumatic heart disease.[5] In this instance, six disease categories were considered. They were normal, mitral insufficiency, mitral stenosis, mitral insufficiency and mitral stenosis, aortic insufficiency, and finally, aortic stenosis. The symptoms considered were pulmonary artery mean pressure, pulmonary artery wedge mean pressure, change in pulmonary artery mean pressure from rest to exercise, total pulmonary resistance and fractional change in cardiac output from rest to exercise. The data values of 171 patients were used to determine the lag normal distributions for each of the symptoms and disease patterns. An attempt was made by the program to diagnose diseases of the left heart chambers when only data from the right heart were measured. This was done to determine if these diseases could be diagnosed using only the right heart data, thus eliminating the necessity of the left heart study when these defects were suspected. However, it was found that aortic stenosis and aortic insufficiency could not be differentiated from the right heart symptoms chosen for the study. There was also a problem in the diagnosis of the mitral insufficiency and mitral stenosis. The patients diagnosed by the cardiologists who had one of the defects were often diagnosed by the program to have both defects. Even though some of the results of this study were discouraging, they did point out the possible effectiveness of the technique as a tool for medical diagnosis.

A second application of this program proved to be highly successful. This application diagnosed polycythemic states using hematological findings on patients.[6] It should be noted here that in the case of normal individuals the normal distribution curve ($\tau=o$) was the most accurate in fitting the data, whereas in those having polycythemia rubra vera, the lag normal distribution curve

TABLE 14–III
COMPARISON OF PERFORMANCE OF PROGRAM WITH HEMATOLOGISTS AND GENERAL PRACTIONERS ON 103 CASES WITH POLYCYTHEMIA RUBRA VERA OR NORMAL

	Correct	*False Positive*	*False Negative*
Program	95	2	3
Hematologists	76	2	22
General Practioners	65	1	34

was the best fit. Table 14–III shows the results of this study on 103 cases with polycythemia rubra vera or normal. As seen from the table, the program was able to diagnose 95 of the cases correctly where experienced hematologists correctly diagnosed 76. The general practitioners were able to diagnose only 65 of the cases correctly. It is of note that three of the symptoms used are interdependent to various degrees—the volume of red packed cells, blood count, and the hemoglobin concentration. To assess the effect this has on the ability of the program to make a correct diagnosis, the three combinations of two red cells parameters were used to the exclusions of the third, in three trial runs. The results of these trials can be seen from Table 14–IV.

A data collection system has been developed which has made it possible to accumulate a large and varied data base on patient's entering two hospitals in Salt Lake City. A computer-based medical record is developed on each patient entering the hospital. Each elective admission is sent through a screening procedure on entry where the following data are collected and entered through a keyboard: vital statistics, such as height, weight, age and sex. A self-administered history is performed by each patient using a

TABLE 14–IV
COMPARISON OF PROGRAM PERFORMANCE WHEN SELECTED MEASUREMENTS OF RED BLOOD CELLS ARE USED ALONG WITH NON-RED CELL PARAMETERS

Measurement Used	*Correct*	*False +*	*False –*
VPRC RBC Hgh	95	2	3
RBC Hgh	93	4	3
RBC VPRC	93	5	2
VPRC Hgh	90	6	4
Hgh	89	3	8

device with which the patient punches a hole in a prepunched card corresponding to each question he wishes to answer in the affirmative. The questionnaire consists of 280 questions. All the answers, plus the patient's identification number, can be punched on a single card.

After this information is entered into the computer through a card reader, a history is printed consisting of a formatted set of positive statements corresponding to the questions answered in the affirmative by the patient. These questions are concerned with a system review, past history, family history, and a history of drug intakes and allergies, and is used primarily for the physician as a screening procedure prior to his investigation of the present illness in depth.

Blood and urine samples are drawn and sent to the laboratory where the blood is analyzed automatically through a 12-channel autoanalyzer and entered directly into the patient's computer-based medical record. Routine hematological and urinalysis data are entered manually from a keyboard in the laboratory.

Spirometry parameters are measured directly from a potentiometer connected to the spirometer into which the patient performs a forced vital capacity. ECG data is analyzed on-line, using a vector approach and the analysis system described above. If either procedure results in an abnormal diagnosis, the test is repeated immediately.

This data, along with other test procedures such as heart catheterization, cardiac output, and other hemadynamic data from intensive care wards, subsequent ECG analysis and follow-up chemistry data are entered automatically into the patient's record. At the time of discharge the patient's diagnosis is coded into his computer record, using a key word approach to generate the codes. Once the diagnosis has been entered, the patient's record is transferred from the active file on magnetic disc to the library tape which may then be searched for statistical purposes.

The first program in the search routine allows the operator to specify any logical combination of diseases to categorize a group of patients. The records satisfying these logical diagnostic codes are then copied from the master tape onto a subtape which can then be further searched to extract the specified data for sub-

sequent analysis. The kind of data and the set of diseases, of course, are determined by the particular diagnostic matrix the operator wishes to generate. For example, he may wish to look at the distribution of total leukocyte counts in patients with leukemia. After completing such a search, a histogram is generated showing this distribution. The operator then enters his first approximation to tau of Equation 5. The computer solves this equation and superimposes the solution over the experimental histogram. If the fit is not optimal from visual inspection the operator may ask for another solution using a different tau value. Once tau is specified, the other two parameters, σ and μ, are determined from the first moment and variance of the experimental data using the relationship described earlier. Thus, a set of lag normal distributions are generated in three parameters for describing the distribution of each variable in each disease and these parameters are stored in a matrix for subsequent use.

The general approach currently being tested is an attempt to first classify the organ system most likely responsible for this set of clinical data obtained at time of admission on each patient. Having once found the most likely organ system involved, a second matrix specifying the nature of the disease involving that organ system may then be tested using some of the same data plus some additional types of tests to achieve this classification.

Of course, most of the patients entering the hospital as elective admissions will already have a primary diagnosis established. The purpose of this screening procedure and the associated diagnostic effort is to find and classify secondary illnesses from which the patient may be suffering and of which the patient's doctor should be aware before initiating treatment for the primary illness. Although some preliminary matrices have been established, critical tests of this approach have not as yet been performed.

Based on the approaches to computerized medical diagnosis presented in this chapter and in other papers, the future holds the exciting promise of continued reliability and improvement of these techniques.

REFERENCES

1. Pryor, T. A.; Russell, R.; Budkin, A. and Price, W. G.: Electrocardiographic interpretation by computer. *Comp Biomed Res*, 2: (No. 6)

Dec. 1969.

2. Warner, H. R.; Toronto, A. F.; Veasy, L. G. and Stephenson, R.: A mathematical approach to medical diagnosis. *JAMA, 177*:177–183.
3. Warner, H. R.; Toronto, A. F. and Veasy, L. G.: Experience with Bayes's Theorem for computer diagnosis of congenital heart disease. *Ann NY Acad Sci, 115*:558–567, July 31, 1964.
4. Reynolds, D. N.: Learning program for computer diagnosis of congenital heart diseases. Master of Science in Computer Science Thesis, Univ. of Utah, 1966.
5. Christensen, J. A.: The lag normal density function applied to computer diagnosis of rheumatic heart diseases. Master of Science in Computer Science Thesis, Univ. of Utah, 1967.
6. Bishop, C. R. and Warner, H. R.: A mathematical approach of medical diagnosis: Application to polycythemic states utilizing clinical findings with values continuously distributed. *Comp Biomed Res, 2*: 486–493, Oct. 1969.